Collagen pentablock copolymers form smectic liquid crystals as precursors for mussel byssus fabrication

Franziska Jehle[1,2], Tobias Priemel[1], Michael Strauss[3], Peter Fratzl[2], Luca Bertinetti*[2,4], Matthew J. Harrington*[1,2]

[1]Dept. of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, Quebec H3A 0B8, Canada

[2]Max Planck Institute of Colloids and Interfaces, Dept. of Biomaterials, Am Mühlenberg 1, 14476 Potsdam, Germany

[3]Dept. of Anatomy and Cell Biology, McGill University, 3640 University Street, Montreal, Quebec H3A 0C7, Canada

[4]BCUBE Center for Molecular Bioengineering, TU Dresden, Tatzberg 41, 01307 Dresden, Germany

Abstract Protein-based biological materials are important role models for the design and fabrication of next generation advanced polymers. Marine mussels (*Mytilus spp.*) fabricate hierarchically structured collagenous fibers known as byssal threads via bottom-up supramolecular assembly of fluid protein precursors. The high degree of structural organization in byssal threads is intimately linked to their exceptional toughness and self-healing capacity. Here, we investigated the hypothesis that multidomain collagen precursor proteins, known as preCols, are stored in secretory vesicles as a colloidal liquid crystal (LC) phase prior to thread self-assembly. Using advanced electron microscopy methods, including scanning TEM and FIB-SEM, we visualized the detailed smectic preCol LC nanostructure in 3D, including various LC defects, confirming this hypothesis and providing quantitative insights into the mesophase structure. In light of these findings, we performed an in-depth comparative analysis of preCol protein sequences from multiple Mytilid species revealing that the smectic organization arises from an evolutionarily conserved ABCBA penta-block co-polymer-like primary structure based on demarcations in hydropathy and charge distribution, as well as terminal pH-responsive domains that trigger fiber formation. These distilled supramolecular assembly principles provide

inspiration and strategies for sustainable assembly of nanostructured polymeric materials for potential applications in engineering and biomedical applications.

Engineering multiscale structural organization into soft materials is vital for advancing fields such as biomedical engineering, bioelectronics, photonics and smart polymers[1-6]. However, assembling nanoengineered molecular constructs into functional macroscale materials presents serious challenges. From this perspective, supramolecular assembly of complex biological materials from protein precursors offers important inspiration. For instance, certain organisms, including spiders, silkworms and mussels, rapidly fabricate hierarchically structured biopolymeric fibers with excellent mechanical properties from fluid protein phases under ambient processing conditions[7, 8]. Mussel byssal threads, which enable attachment in wave-swept seashore habitats (Fig. 1a), are an established model system for inspiring advanced soft materials and adhesives[9-11]. A growing literature suggests that mussels employ condensed fluid protein precursor phases to guide self-assembly of hierarchical structure in byssal threads[8, 12]. Specifically, mussels appear to utilize coacervates (i.e. liquid-liquid phase separations (LLPS)) to produce a microporous underwater adhesive[12, 13], and to deploy co-existing condensed LLPS phases comprised of several functionally distinct proteins to produce a hard, yet extensible composite coating[14, 15]. Of particular relevance to the current study, mussels also seem to use liquid crystalline (LC) phases to build the tough and self-healing collagenous core of the byssal threads[16-18]. Despite recent progress on these topics, many questions are unanswered. Here, we utilized advanced electron microscopic imaging to further investigate the hypothesis that mussels utilize a LC collagen precursor phase to initiate rapid assembly of the hierarchically structured self-healing core[16, 18, 19].

The distal region of the thread core (closest to the substrate) is comprised almost entirely of an elaborate hierarchical organization of collagenous proteins known as preCols[20-23] (Fig. 1f-i). There are three preCol variants, each of which is a multidomain chimeric protein consisting of a central rigid triple helical collagen domain[20]. At both ends of the collagen domain are flanking domains that vary in sequence between the three variants resembling dragline silks beta sheet sequences (preCol-D), flagelliform silk (preCol-NG) or unstructured elastin motifs (preCol-P)[21-23]. Located at

the N- and C-terminal ends of all preCols are domains enriched in the amino acid histidine, which are known to form metal coordination complexes that contribute to fiber toughness and self-healing as reversible sacrificial bonds (Fig. 1g)[20, 24]. Detailed studies utilizing X-ray diffraction and various spectroscopy methods have revealed the multiscale organization of preCols within the threads are critical for achieving their exceptional tensile properties[24-29]. In particular, evidence suggests that preCol triple helices are organized into 6+1 hexagonal bundles that are the functional subunits for forming the thread (Fig. 1h) and that preCol bundles are further organized into a semi-crystalline framework with quasi hexagonal lateral spacing and a highly defined axial staggering of 13 nm (Fig. 1i)[18, 26]. *In situ* studies coupling mechanical testing with X-ray and spectroscopic measurements reveal the crucial role of hierarchical structure in determining thread mechanics[24-27, 29], but how is such a complex multiscale structure formed in the first place?

A mature mussel byssus can consist of more than 100 threads with lengths of several centimeters and diameters of 100-250 μm, each of which is fashioned individually by the mussel. Thread formation is a regulated exocrine secretion process lasting 3-5 minutes in which the protein contents of secretory vesicles stockpiled in specialized glands are released into a narrow groove running along the mussel foot – the organ that forms the byssus (Fig. 1a-d)[8, 12, 30, 31]. Assembly of concentrated protein precursors has been proposed to be triggered in part by the pH transition from the presumed acidic conditions within the secretory vesicles to the basic pH of seawater[17, 18]. The core gland (alternatively called white gland or collagen gland[8, 30, 32]) produces and stockpiles prolate ellipsoid secretory vesicles (~2 μm long axis) that are filled with preCols in a concentrated fluid phase[16] (Fig. 1a-c). Based on strong birefringence of the vesicles[8, 31] and their characteristic layered banding pattern in TEM images[32] (Fig. 1c,d), preCols were hypothesized to be stored as a smectic LC phase[16, 18, 19]. The smectic phase is proposed to be a means of pre-organizing the preCols in an ordered fluid state that can be rapidly mobilized and assembled into the highly organized thread hierarchical structure (Fig. 1e, i)[8, 16, 18]. However, the preCol LC phase has been exceedingly challenging to characterize experimentally at the nanoscale, especially in 3D – although this information is crucial for understanding the byssus assembly process. Moreover, the specific physicochemical features that favor smectic organization by preCols have

not been clearly identified. Answering these open questions is the key for understanding and mimicking the remarkable transition from fluid precursor to hierarchically structured fiber.

In the present study, we employed advanced electron microscopy including focused ion beam scanning electron microscopy (FIB-SEM) to investigate the 3D nanostructure of the LC protein phase within core secretory vesicles with unprecedented resolution. The structural data were analyzed in light of existing biochemical data, revealing new insights into the forces driving smectic LC self-organization, as well as pH-triggered self-assembly into high-performance polymer fibers. Based on their excellent material properties and collagenous composition of byssal threads, these new insights hold strong potential for inspiring new avenues for sustainably producing high-performance polymeric and bio-polymeric materials for both engineering and biomedical applications.
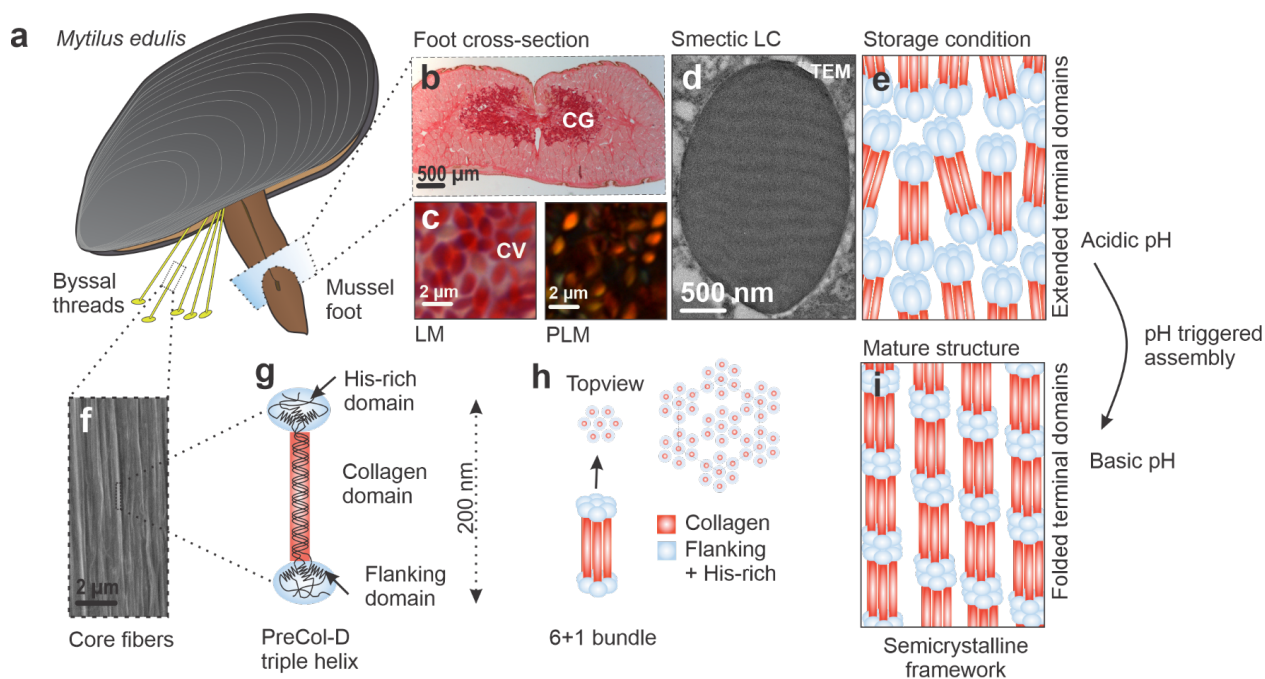

Fig. 1 – Overview of mussel byssus hierarchical structure and assembly. a) Mussels fabricate tough adhesive attachment fibers called byssal threads with an organ called the foot. b) Foot cross-section stained with Sirius red showing the location of the core gland (CG). c) Light microscopy (LM) and polarized light microscopy (PLM) images from a magnified region of the core gland showing Sirius red stained and birefringent ellipsoid core vesicles (CV). d) TEM micrograph of a core vesicle showing characteristic banding pattern. e) Proposed liquid

crystalline organization of preCols inside core vesicles. f) SEM image of distal byssal thread core showing fibrous morphology. g) Generalized schematic of preCol triple helical structure showing different structural domains. h) PreCol triple helices self-organize into 6+1 bundles with hexagonal lateral packing, which are further arranged in a semicrystalline structure in native threads (i) with 13 nm axial stagger between adjacent fibrils. Assembly from the storage LC phase to the structured fiber is proposed to be pH-triggered (e and i).

Results

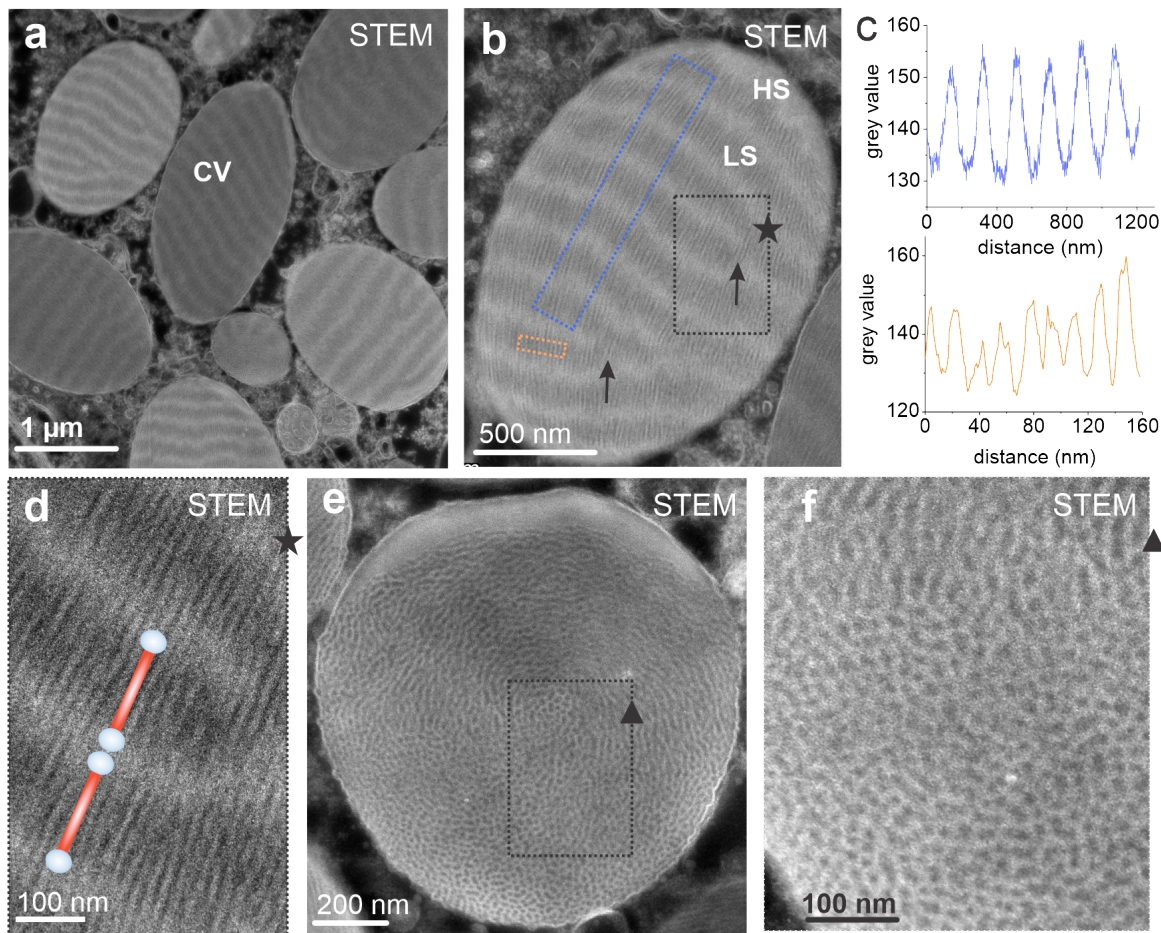### Scanning TEM investigation of secretory vesicles

Fig. 2 – STEM imaging of core vesicle LC phases. a) STEM micrograph of osmium-stained sections from the core gland showing numerous banded core vesicles (CV). b) Higher magnification image of a single core vesicle (cut parallel to long vesicle axis) showing the heavy staining (HS) and light staining (LS) regions of the banding pattern. Arrows identify defects in the banding pattern. c)

Intensity plots of the banding periodicity (top, blue) and lateral filament periodicity (bottom, orange) taken from the blue and orange boxes in (b), respectively. d) Higher magnification image of the region in the starred box in (b) highlighting the filamentous nature of the LS band. PreCol schematic is overlapped based on the hypothesis that the filaments represent the collagen domain of the preCols. e) STEM micrograph of a transverse section of a core vesicle (cut perpendicular to long vesicle axis). f) Magnified image of the region in the box with a triangle in (e) show cross-section and packing of filaments in the LS band.

Fig. 1d shows a transmission electron microscopy (TEM) image of a core vesicle from a thin section of a fixed and stained mussel foot revealing the characteristic banding pattern observed in previous studies[32-34] consisting of alternating dark heavy staining (HS) and light staining (LS) layers. Scanning TEM (STEM) images (Fig. 2) provide a higher degree of contrast than standard TEM, enabling imaging of the layer structure at higher resolution (*n.b.* contrast is inverted – HS is bright, LS is dark). Closer examination reveals that the HS layer appears to be less extended than the LS layer and shows very weak contrast, indicating a lack of internal structure; whereas the LS layer exhibits a fibrillar texture with packed filaments aligned perpendicular to the layers (Fig. 2a-d). Transverse STEM sections further reveal that the dark (LS) filaments are evenly spaced with a lighter (HS) material separating the filaments (Fig. 2e-f). FFT analysis of the longitudinal and transverse section both suggest a center-to-center distance of ~19 nm between the dark filaments, consistent with spacings predicted between preCol bundles in native threads based on AFM and SAXS measurements[18, 26]. Although challenging to accurately measure in these 2D images, the diameter of the dark filament is about half the center to center spacing – consistent with the expected diameter of the 6+1 preCol bundle consisting of 7 collagen triple helices (Fig. 1h, 2f) [18, 19, 26, 32]. Accordingly, the nonfibrillar HS layer likely represents the overlapping ends of two layers of preCol bundles consisting of flanking domains and HRDs (Fig. 1e, g, 2d). The layered molecular arrangement is consistent with a smectic colloidal LC arrangement, and layer disruptions indicating what appear to be LC edge disclinations are observed in numerous vesicles (Fig. 2a-b). It is worth noting that we do not observe a bent core shape in the filaments of the preCol smectic phase using STEM that was previously reported in native threads with AFM; however, this may appear only after thread formation[18].

FIB-SEM reconstruction of vesicle LC phase

While STEM image analysis of vesicles suggests a smectic LC phase, sectioning and imaging of complex 3D phase structures in 2D leads to distortions, making accurate quantitative analysis of layer structure and defects challenging. In order to investigate the 3-dimensional LC textures within core vesicles, we utilized FIB-SEM followed by 3D reconstruction of image stacks from small volumes within core gland samples (Fig. 3a, Movie S1). Fig. 3b shows a single SEM image from the image stack showing that the ellipsoidal core vesicles are tightly packed within the cells of the gland tissue consistent with histological sections (Fig. 1b-d). Vesicles cut nearly parallel to the long axis exhibit a clear banding pattern consistent with the TEM images, arising from differential staining; however, due to the lower spatial resolution of FIB-SEM, it is not possible to resolve the individual preCol filaments. It is worth noting here that the lighter bands indicate higher electron density, which is the same as STEM contrast – thus, the dark layers contain the filaments observed in STEM.

To visualize these banding patterns in 3D, voxels associated with the core vesicles within the image stacks were computationally reconstructed into 3D objects. Fig. 3c shows around 25 reconstructed vesicles as they are organized in the gland in which only the HS phase is visualized Details of individual vesicles are shown in Fig. 3e-g and supporting movies S2-S6. Quantitative 3D image analysis of vesicles (n = 56 vesicles from two different mussel feet) revealed a prolate ellipsoidal shape with a long axis of 2.3 ± 0.3 µm and an aspect ratio of 2.1 ± 0.3 (Fig. S1), consistent with previous light microscopy measurements[16]. Within the vesicles, layer spacing was 180 ± 7 nm (including both the HS and LS layers) [18, 25]. In contrast to well-defined layer spacing, there was significant variability in the angle the layers make relative to the vesicle long axis with most ranging from 0° to 45° with an average of 18° ± 11° (Fig. 3c, Fig. S1). In vesicles with smectic layers oriented nearly perpendicular to the long axis, the layers are typically separate and discontinuous, whereas layers oriented at larger angles typically interact with neighboring layers, producing distinctive defects including screw disclinations (Fig 3d-g, Movies S2-S6). These variations in layer arrangement can be visualized by plotting the angle of the layer normal in spherical coordinate graphs (Fig. 3d-g). These observations clearly underscore the colloidal smectic LC nature of the phases. Indeed, such variability is to be expected in an LC phase and is

likely influenced and frustrated by confinement within a membrane[16], as previously observed with cholesteric cellulosic LCs confined in microfluidic droplets[35]. Most importantly, smectic layering is a conserved structural feature of all vesicles with a thickness that is in the range of the predicted length of the preCols[18, 25].
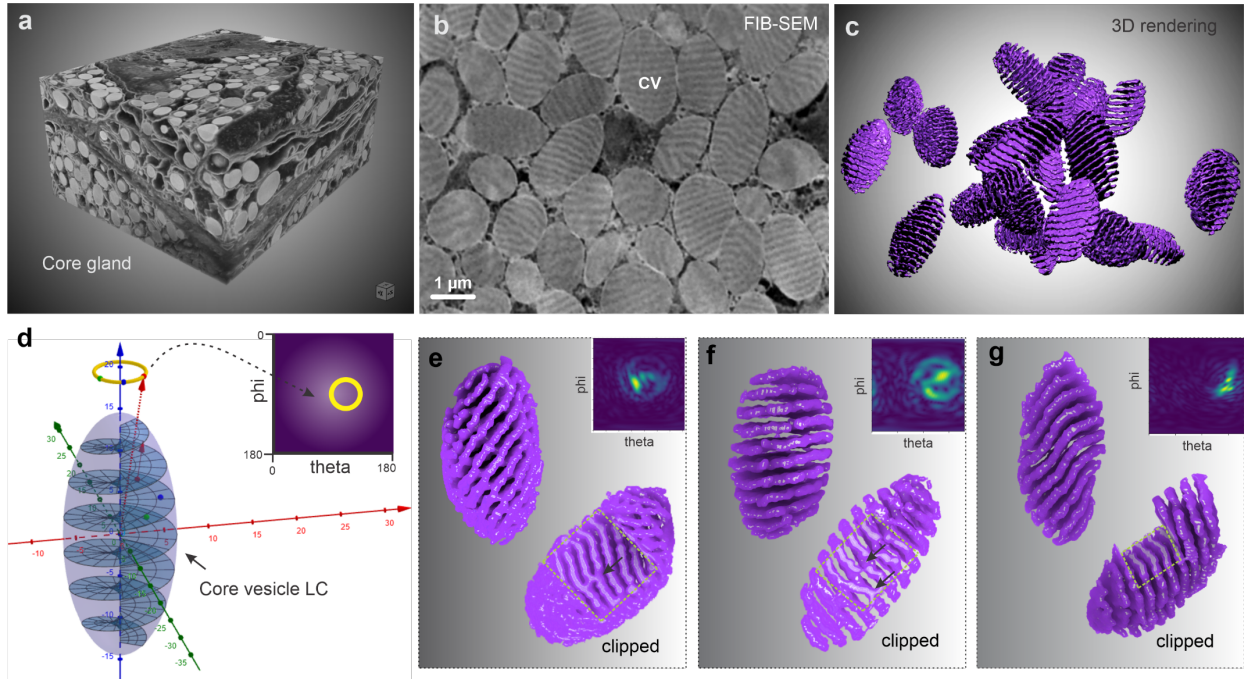

Fig. 3 – FIB-SEM imaging and tomographic reconstruction of core vesicle LC phases. a) 3D reconstruction of FIB-SEM image stack acquired from a small volume of the core gland showing cellular storage of core vesicles (CV). b) Single image from the FIB-SEM image stack showing banded core vesicles. c) 3D rendering of ~25 individual core vesicles from the image stack volume, showing only the HS layer of the banded texture in the core vesicles. d) Schematic showing origin of spherical coordinate graphs characterizing the orientation and helicity of the layers in the vesicles. The tip of a vector drawn perpendicular to the HS layer planes is traced in the phi-theta plane. e-g) 3D reconstructions of 3 individual vesicles from (c) showing different layer morphologies and disclinations. Each vesicle is viewed from two different vantage points and in the lower vantage point, the inner texture is exposed to highlight the internal layer structure. The spherical coordinate graph for each vesicle from analysis based on the schematic in (d) is shown in the top right corner highlighting the differences in layer arrangements.

Sequence analysis

The EM investigations provide new insights into the smectic preCol packing from the nanometer to micron-scale, suggesting a key role of the different preCol domains in guiding self-organization. Here, we reexamine the primary sequence of the preCols for insights into physicochemical forces driving smectic self-organization[19, 20, 36, 37]. Figure 4 provides a visualization of the hydropathy and charge distribution along the length of the three preCol variants D, NG and P from *Mytilus edulis* under the putative storage conditions (pH 5) and the assembly conditions (pH 8)[17, 18]. Plots reveal that despite significant differences in the sequence motifs between the variants, all preCols exhibit a highly consistent symmetric pentablock co-polymer-like pattern of hydropathy and charge with an ABCBA block organization (Fig. 4c). More specifically, these blocks consist of a central charged polar region making up approximately 50% of the total sequence (C block), two adjacent non-polar regions devoid of charged residues that each range between 11-17% of the sequence (B blocks) and two terminal polar domains making up between 9-12% of the total sequence (A blocks). Importantly, the terminal domains are highly positively charged under low pH storage conditions and largely uncharged under basic seawater conditions arising from the high local molar concentration of histidine (pK_a = 6.5) (Fig. 4c-d, Table S1). Analogous plots of preCol sequences from two related species *Mytilus galloprovincialis*[37] and *Mytilus californianus*[36] are highly similar to *M. edulis* also possessing the characteristic pentablock structure, although the overall length of the proteins and sequences vary, indicating that there is a strong evolutionary pressure to conserve this pentablock pattern of charge and hydropathy across species and variants (Fig. S2 and S3, Table S1). This is extremely relevant to the observed organization of the preCol LC phase as these block-like variations likely create barriers that prevent sliding of the different domains past one another in the LC phase, maintaining smectic register, as observed in synthetic block co-polymers[38-40].
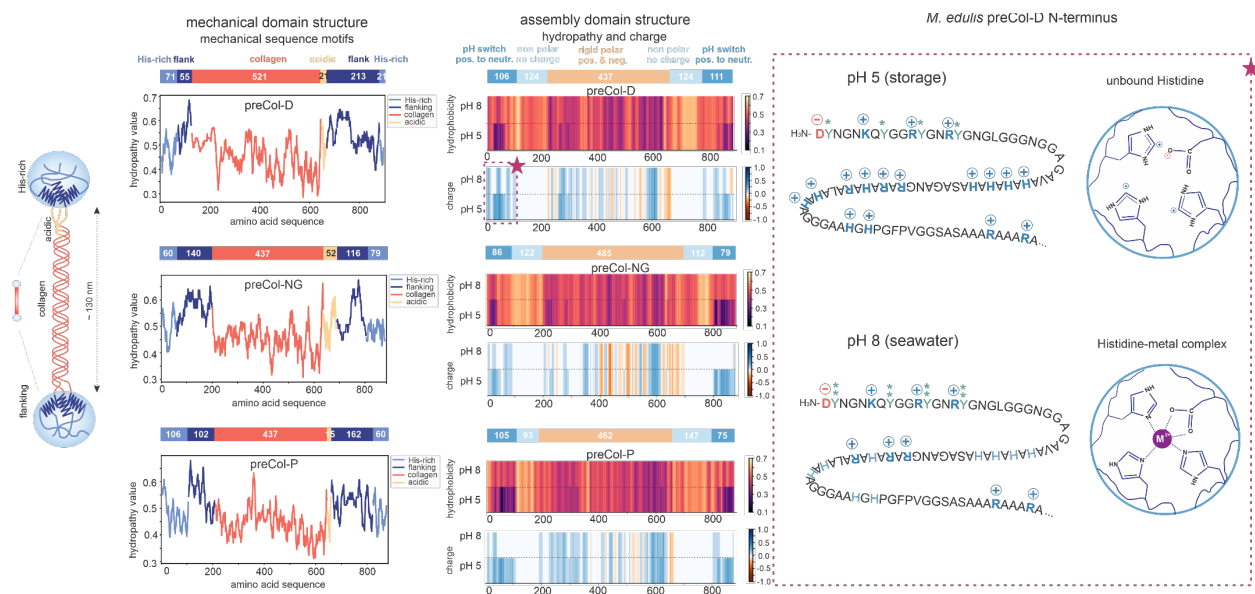
Fig. 4 – Charge and hydropathy profiles of mussel preCol block co-polymer proteins. a) Generalized schematic of a preCol triple helix showing the previously assigned domains based on similarity to known protein structural motifs. b) Hydropathy plot of the three preCol variants from *M. edulis* with mechanical domains identified. Higher values indicate more hydrophobic character and lower values indicate more hydrophilic. c) Heatmap plots of hydropathy and charge distribution for all three preCols variants from *M. edulis* under storage (pH 5) and seawater (pH 8) conditions. Clear and consistent variations can be seen that are highly similar between all three variants. d) Sequence and charge distribution of the N-terminus of preCol-D (starred box in (c)) under storage and seawater pH conditions, showing deprotonation of the histidine residues. Similar behavior occurs at the N- and C-termini of all preCol variants, enabling the formation of His-metal coordination bonds and cross-linking of the LC phase into a tough fiber. DOPA residues (Y*) are also well conserved at the termini, likely becoming oxidized to their quinone form (Y**) at high pH, leading to subsequent oxidative cross-linking as previously proposed[41].

Interestingly, in preCol-D these physicochemically distinct blocks do not correspond precisely to the previously assigned domain boundaries, which were based on homology to mechanically relevant sequence motifs (Fig. 1G, 4a-c, Table S1)[20]. For example, at its N-terminal end, the collagen domain of preCol-D is highly hydrophobic and completely devoid of charged residues, while the remainder is very hydrophilic, containing both positively and negatively charged amino acids. Similarly, the C-terminal flanking domain of preCol-D contains a more hydrophilic charged

portion on the C-terminal side and an uncharged portion on the N-terminal side. This suggests that, at least for preCol-D, the variation in hydropathy and charge, which we posit control LC organization and assembly, have been determined by different selective pressures than the motifs that determine mechanical properties in the threads – even though they constitute the same sequence.

Given the number of amino acids in the collagen domains of the different *M. edulis* preCol sequences[20] (Fig. 4a) and the previously measured rise per residue of 0.29 nm for the collagen triple helix in byssal threads[25], we estimate that the length of the collagen domains will be 122 nm for preCol-NG and preCol-P and will be 151 nm for preCol-D. Thus, we assume that additional length of the smectic spacing (~180 nm) observed in FIB-SEM (Fig. 3, S1C) constitutes the combined length of the overlapping N- and C-terminal flanking and His-rich domains in the smectic phase (~30-60 nm). Notably, SAXS analysis of native threads predicts that the flanking + His-rich domains should be packed in a length of less than 13 nm[26], which is consistent with the dominant cross-beta sheet structure observed with WAXD, which is a highly efficient means of packing long protein sequences in very small volumes[27]. Thus, we conclude that within the smectic LC phase in the core vesicles, the preCol flanking and His-rich domains are unfolded and at least partially extended, and that they are highly folded within the thread. This is supported by the fact that TEM of native threads does not show a smectic banding pattern[42]. Moreover, it offers support for the hypothesis that assembly may be driven by a pH-triggered protein folding process initiated at the terminal domains of the preCols.

Discussion

Our EM investigations provide unequivocal evidence substantiating the hypothesis that the fluid protein phase within the core secretory vesicles of marine mussels is a smectic LC[18, 19]. It is assumed that the preorganization of the preCol bundles is crucial for rapid self-assembly of the thread hierarchical structure, which determines the high toughness and self-healing response [10, 24, 27]. Essentially, within the smectic phase, the preCol precursors are already organized orientationally and positionally in a manner similar to the final structure, especially with regards to ensuring that flanking domains and His-rich domains of adjacent preCol bundles are in close

contact. Yet, as a fluid phase, the smectic LC is also malleable and able to be molded and processed quickly into a macroscopic fiber. This is analogous to the role of nematic LCs in the processing of high-performance polymers such as Kevlar[43]. However, unlike Kevlar, which is produced by dissolving petroleum-based precursors in pure sulfuric acid at 80°C to achieve the LC phase[43], byssus formation occurs in the ocean under ambient conditions, making it an inherently more environmentally friendly process. Moreover, the rapid self-assembly of these collagen-based byssus fibers in just minutes[8, 30] is truly impressive considering the extended length of time required by the human body for cell-dependent production and healing of collagenous tissues, such as tendon and bone[44, 45]. In this light, the use of stimuli-responsive smectic mesogens as precursors for material assembly has great potential in the realm of both sustainable polymers and biomedical scaffolds.

A number of biomolecules are known to form nematic and cholesteric LC phases under both *in vivo* and *in vitro* conditions, including collagen, silk spidroin, amyloid fibers, chitin, cellulose and even DNA[46-50] . However, aside from preCols, we are unaware of other examples of biomolecules that naturally form smectic phases *in vivo* as an evolved feature for material assembly (although rod-like tobacco mosaic virus capsids can be induced to form smectic phases at artificially high concentrations[51]). This is perhaps not surprising as smectic phases possess a significantly higher degree of order than nematics, and thus, the physical and chemical parameters required for forming a smectic phase are more stringent[52]. Our comparative analysis of nine different preCol sequences from three different mussel species reveals that the tendency of preCols to naturally self-organize into smectic phases likely arises from an evolutionarily conserved symmetric ABCBA penta-block co-polymer-like pattern of charge and hydropathy encoded into the primary sequence of all preCols. Interestingly, our sequence analysis suggests that there may be separate biochemical control of assembly and mechanics and thus, multifunctionality of the protein sequence in the fluid vs. fiber state. For example, from a mechanical perspective, the collagen domain of preCol-D characterized by a characteristic Gly-X-Y sequence provides rigidity to the native thread via triple helix formation[25], whereas the silk-like domains form cross beta sheet structures that contribute hidden length and reversible extensibility[27-29]. Yet, both domains are also subdivided into distinct non-polar and polar regions that contribute to the identified

pentablock co-polymer domains that we propose drive smectic phase formation under the highly concentrated conditions in the vesicles.

In addition to these more chemical driving forces, LC mesogens (i.e. molecules that form LC mesophases) often possess distinctive physical features that contribute to LC formation. For example, a rigid rod-like core possessing a persistence length greater than the contour length is a common characteristic of both synthetic and biological mesogens that contribute to the anisotropy of the resulting mesophase [47, 52]. The persistence length of a type I collagen triple helix has been predicted to be in the range of 80-100 nm[53]; however, the preCol mesogen consists of 7 triple helices (21 polypeptide chains) arranged in a 6+1 helical bundle (Fig. 1h), which will certainly increase the bending rigidity and consequently, the persistence length. In addition to a rigid core, many synthetic mesogens possess flexible chain-like elements at their termini, which contribute to their fluid-like properties under certain conditions[52]. In this light, the unfolded, flexible flanking domains at both ends may provide further impetus to favor a fluid smectic phase under acidic storage conditions, especially as they will carry a repulsive positive charge[54]. Finally, smectic phases are favored by synthetic rod-like mesogens that possess a symmetric structure along the length that favors packing of mesogens into lamellar structure, as observed in all preCols[52].

The symmetric ABCBA block co-polymer structure of the preCols consisting of a rigid collagen core and flexible ends fulfills each of the above requirements for forming a fluid smectic phase under storage conditions, but how is this phase converted suddenly to a hierarchically structured fiber? Again, there appear to be both physical and chemical mechanisms at play. From a physical perspective, previous studies show that core vesicles and purified preCols can be induced to form mesoscale structured fibers *in vitro* via shear or tensile manipulation [8, 16, 17], suggesting an important role of mechanical forces in alignment and assembly of preCols. Similar shear-induced alignment and formation of smectic lamellar phases was previously observed with synthetic penta-block co-polymer molecules[40]. From a more chemical perspective, our sequence analysis (Fig. 4) supports the hypotheses that a histidine-dependent pH-switch going from the slightly acidic vesicle conditions to seawater pH may initiate the LC to fiber transition[16-19]. This pH transition will lead to mass deprotonation of the histidine side chains at the preCol terminal

domains, removing repulsive terminal interactions at the flexible ends, enabling protein folding and also formation of histidine-metal protein interactions that contribute as sacrificial bonds in toughening and self-healing (Fig. 4c-d)[24]. Indeed, it was shown that synthetic peptides based on preCol His-rich motifs undergo a rapid pH-dependent conformational transition from random coil to cross-beta crystallites mechanically reinforced with metal ions[55, 56] that could be used to build hierarchically structured, mechanically tunable materials[57, 58]. Recent reports further indicate that conserved DOPA catechol residues at the very termini of preCols may initiate a secondary slower redox-based covalent cross-linking after initial supramolecular assembly based on histidine deprotonation (Fig. 4d)[41]. In this light, a key role of the smectic phase may be to ensure that His-rich domains of neighboring preCol bundles are in contact at the moment of assembly, to favor proper alignment for cross-linking of adjacent preCol bundles. However, after formation, the dense packing of the His-metal coordination network and beta sheet structure produces the previously proposed double network of hidden length and sacrificial bonds that provides the self-healing and tough behavior[27].

Conclusion

The ability of preCols to form smectic LC phases that can be easily processed into hierarchically structured tough and self-healing fibers is encoded in the primary protein sequence in the form of an ABCBA penta-block structure. The physicochemical assembly mechanisms elucidated here are highly relevant and adaptable to current efforts to produce high-performance polymeric materials and supramolecular nanomaterials under environmentally friendly conditions. These design principles could conceivably be engineered into the structure of synthetic polymers. For example, block co-polymer molecules have already been observed to self-organize into smectic-like phases very similar to what is observed in the mussel[38-40]. However, building hierarchically structured bulk materials from such precursor phases under aqueous conditions requires getting the balance of hydropathy and charge in the blocks, as well as the balance of mechanical rigidity and flexibility correct. Moreover, the preCols provide an attractive pH-based assembly trigger that can be encoded into blocks. This aspect has already demonstrated in simpler mussel-inspired self-healing materials based on imidazole chemistry[59-61]. In addition to their more technical potential, these extracted concepts are also extremely relevant to ongoing efforts to produce

injectable biomedical scaffolds[62, 63], especially given the collagenous nature of the building blocks.

Methods

Materials. Blue mussels (*Mytilus edulis*) were purchased and maintained at ~14 °C in an aquarium with artificial saltwater. Investigations were performed on the foot organ of adult mussels removed with a scalpel. We have complied with all relevant ethical regulations for testing and research of *Mytilus edulis*.

Chemical fixation and embedding. Dissected feet were carefully rinsed with cold water, blotted with a paper towel to remove mucus and pre-fixed for 30 min at 4 °C in 3 % glutaraldehyde, 1.5 % paraformaldehyde, 650 mM sucrose in 0.1 M cacodylate buffer pH 7.2. The foot tissue was then cut into thin cross-sections comprising the groove and part of the gland tissue and then fixed for 2 h at 4 °C in the same buffer as above. Fixed samples were rinsed 5× with 0.1 M cacodylate buffer, pH 7.2 at 4 °C and post-fixed with 1 % $OsO4$ for 1 h at 4 °C. Samples were rinsed again in 0.1 M cacodylate buffer pH 7.2 (3 × 5 min at 4 °C), followed by series dehydration in acetone (50 %, 70 %, 90 %, 3 x 100 %) for 10 min each step at RT. Dehydrated samples were embedded in Epoxy (Epon 812 substitute, Sigma-Aldrich, # 45359) for TEM and FIB-SEM and polymerized at 70 °C for at least 48 h. Ultrathin sections of 100 nm for STEM investigations were prepared using an ultramicrotome and mounted on carbon coated Cu grids (200 mesh).

Scanning Transmission Electron Microscopy. Transmission electron microscopy (TEM) and scanning transmission electron microscopy (STEM) was performed with a Thermo Scientific Talos F200X G2 S/TEM equipped with a Ceta 16M CMOS Camera, operated at 200 kV acceleration voltage. STEM mode was used for High Angle Annular Dark Field (HAADF) imaging at magnifications of 16,500×, 46,000×, 66,000× and 130,000×.

FIB-SEM. Resin blocks containing foot tissue samples (n = 2) were polished in order to expose the tissue at the block surface. Samples were sputter-coated with three Carbon layers (~ 5 nm each) and one platinum layer (~ 5-10 nm) or with a 10 nm gold layer and transferred to the Zeiss Crossbeam 540 (Carl Zeiss Microscopy GmbH, Germany). At the region of interest, a trench for SEM imaging was milled into the sample surface using a current of 30-65 nA FIB at 30 kV acceleration voltage. The resulting cross-section was finely polished using the 1.5 nA FIB probe at 30 kV. Thin slices of samples were removed in a serial manner by FIB milling (300 pA, 30 kV, slice thickness 17.5 nm (n1); 700 pA, 30 kV, slice thickness 31.5 nm (n2)). After each milling step, the specimen was imaged by SEM (acceleration voltage = 2 kV) using the secondary and backscattered electron detector (a grid potential of 1.5 kV was set for the EsB detector). The image resolution was 2048 × 1536 pixels with a lateral image pixel size of 18.01 nm (n1) or 26.5 nm (n2). Images were recorded using line averaging (N = 4) and a dwell time of 200 or 50 ns.

FIB-SEM Data Processing. The resulting secondary electron images were processed using the SPYDER3 (Scientific Python Development Environment) (Python 3.6) software. Custom-written python scripts were developed and provided by Luca Bertinetti. Images were automatically aligned using enhanced correlation coefficient alignment. Denoising was performed by applying noise2void in 3D mode[64]. Sauvola's local thresholding computation was applied (as implemented in the scikit-image library, v. 0.17.0) with a block size of 9, a k value of 0.002 (n=1) and 0.01 (n=2) and r value of 4[65]. As the thresholded images contained regions resulting from statistical noise in the image, only thresholded regions containing minimum 10 pixels were selected. Images were median 3D filtered with Fiji and x, y, z radii were set to 1.5. Segmentation of cuticle vesicles was performed using the Amira 2020.1 software (Thermo Fisher Scientific, USA). Shapes of 30 individual vesicles per sample were segmented manually from processed secondary electron images using the brush tool. The HS phase of each vesicle was automatically segmented using the Magic Wand tool by using the local threshold computed and median3D filtered image stacks. 3D visualization of the HS phase was realized by creating meshes of segmented structures in Dragonfly 4.1 (Object Research Systems, Montreal, Canada) and by using Drishti for clipping. Movies of the 3D rendered structures were prepared with the Movie Maker tool in Dragonfly 4.1.

Charge and Hydrophobicity Plots. Hydropathy plots were generated by using the Black scale[66], since it provides values for the post-translational modified amino acid hydroxyproline and can distinguish between neutral and positive charged histidine. For the charge distribution plots, negatively charged amino acids were assigned a value of -1, positively charged amino acids a value of +1 and all other amino acids a value of 0. For both the hydrophobicity plots and the charge distribution plots, we assumed that all histidine residues are positively charged at pH 5, whereas at pH 8 all histidine residues are neutral. A window size of 13 was used for all plots, meaning that the value for a residue at a specific position was calculated by averaging its value together with the values of the 6 previous and 6 following residues. For visualizing the data, we plotted the data as 1D heatmaps using Python 3.6, where values are represented as colors on a color scale.

Acknowledgements

References

1.   Meseck, G. R.; Terpstra, A. S.; MacLachlan, M. J., Liquid crystal templating of nanomaterials with nature's toolbox. *Current Opinion in Colloid & Interface Science* **2017,** *29*, 9-20.

2.   Sai, H.; Tan, K. W.; Hur, K.; Asenath-Smith, E.; Hovden, R.; Jiang, Y.; Riccio, M.; Muller, D. A.; Elser, V.; Estroff, L. A.; Gruner, S. M.; Wiesner, U., Hierarchical porous polymer scaffolds from block copolymers. *Science* **2013,** *341*, 530-534.

3.   Lutolf, M. P.; Hubbell, J. A., Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nat. Biotechnol.* **2005,** *23*, 47-55.

4.   Lutz, J. F.; Lehn, J. M.; Meijer, E. W.; Matyjaszewski, K., From precision polymers to complex materials and systems. *Nat. Rev. Mater.* **2016,** *1*, 16024.

5.   Webber, M. J.; Appel, E. A.; Meijer, E. W.; Langer, R., Supramolecular biomaterials. *Nat. Mater.* **2016,** *15*, 13-26.

6.   Lehn, J. M., Perspectives in chemistry - Steps towards complex matter. *Angew. Chem. Intl. Ed.* **2013,** *52*, 2836-2850.

7.   Heim, M.; Keerl, D.; Scheibel, T., Spider silk: From soluble protein to extraordinary fiber. *Angewandte Chemie - International Edition* **2009,** *48*, 3584-3596.

8.   Priemel, T.; Degtyar, E.; Dean, M. N.; Harrington, M. J., Rapid self-assembly of complex biomolecular architectures during mussel byssus biofabrication. *Nature Communications* **2017,** *8*, 1-12.

9.   Harrington, M. J.; Jehle, F.; Priemel, T., Mussel byssus structure-function and fabrication as inspiration for biotechnological production of advanced materials. *Biotechnology Journal* **2018,** *13*, 1800133.

10.   Lee, B. P.; Messersmith, P. B.; Israelachvili, J. N.; Waite, J. H., Mussel-inspired adhesives and coatings. *Annu. Rev. Mater. Res.* **2011,** *41*, 99-132.

11.   Li, L.; Smitthipong, W.; Zeng, H., Mussel-inspired hydrogels for biomedical and environmental applications. *Polym. Chem.* **2015,** *6*, 353-358.

12.   Waite, J. H., Mussel adhesion - Essential footwork. *Journal of Experimental Biology* **2017,** *220*, 517-530.

13.   Valois, E.; Mirshafian, R.; Waite, J. H., Phase-dependent redox insulation in mussel adhesion. *Science Advances* **2020,** *6*, eaaz6486.

14.   Jehle, F.; Macías-Sánchez, E.; Fratzl, P.; Bertinetti, L.; Harrington, M. J., Hierarchically-structured metalloprotein composite coatings biofabricated from co-existing condensed liquid phases. *Nature Communications* **2020,** *11*, 862.

15.   Monnier, C. A.; DeMartini, D. G.; Waite, J. H., Intertidal exposure favors the soft-studded armor of adaptive mussel coatings. *Nat. Commun.* **2018,** *9*, 3424.

16.   Renner-Rao, M.; Clark, M.; Harrington, M. J., Fiber formation from liquid crystalline collagen vesicles isolated from mussels. *Langmuir* **2019**, 15992–16001.

17.   Harrington, M. J.; Waite, J. H., pH-dependent locking of giant mesogens in fibers drawn from mussel byssal collagens. *Biomacromolecules* **2008,** *9*, 1480-1486.

18.   Hassenkam, T.; Gutsmann, T.; Hansma, P.; Sagert, J.; Waite, J. H., Giant bent-core mesogens in the thread forming process of marine mussels. *Biomacromolecules* **2004,** *5*, 1351-1354.

19.   Waite, J. H.; Vaccaro, E.; Sun, C.; Lucas, J. M., Elastomeric gradients: a hedge against stress concentration in marine holdfasts? *Philos Trans R Soc Lond B Biol Sci* **2002,** *357*, 143-153.

20.   Waite, J. H.; Qin, X. X.; Coyne, K. J., The peculiar collagens of mussel byssus. *Matrix Biology* **1998,** *17*, 93-106.

21.   Coyne, K. J.; Qin, X.-X.; Waite, J. H., Extensible collagen in mussel byssus: a natural block copolymer. *Science* **1997,** *277*, 1830-1832.

22.   Qin, X.-X.; Waite, J. H., A potential mediator of collagenous block copolymer gradients in mussel byssal threads. *Proceedings of the National Academy of Sciences* **1998,** *95*, 10517-10522.

23.   Qin, X. X.; Coyne, K. J.; Waite, J. H., Tough tendons - mussel byssus has collagen with silk-like domains. *Journal Of Biological Chemistry* **1997,** *272*, 32623-32627.

24.   Schmitt, C. N. Z.; Politi, Y.; Reinecke, A.; Harrington, M. J., Role of sacrificial protein-metal bond exchange in mussel byssal thread self-healing. *Biomacromolecules* **2015,** *16*, 2852-2861.

25.   Harrington, M. J.; Gupta, H. S.; Fratzl, P.; Waite, J. H., Collagen insulated from tensile damage by domains that unfold reversibly: In situ X-ray investigation of mechanical yield and damage repair in the mussel byssus. *Journal of Structural Biology* **2009,** *167*, 47-54.

26.   Krauss, S.; Metzger, T. H.; Fratzl, P.; Harrington, M. J., Self-repair of a biological fiber guided by an ordered elastic framework. *Biomacromolecules* **2013,** *14*, 1520-1528.

27.   Reinecke, A.; Bertinetti, L.; Fratzl, P.; Harrington, M. J., Cooperative behavior of a sacrificial bond network and elastic framework in providing self-healing capacity in mussel byssal threads. *Journal of Structural Biology* **2016,** *196*, 329-339.

28.   Arnold, A. A.; Byette, F.; Seguin-Heine, M. O.; LeBlanc, A.; Sleno, L.; Tremblay, R.; Pellerin, C.; Marcotte, I., Solid-state NMR structure determination of whole anchoring threads from the blue mussel *Mytilus edulis*. *Biomacromolecules* **2013,** *14*, 132-141.

29.   Hagenau, A.; Papadopoulos, P.; Kremer, F.; Scheibel, T., Mussel collagen molecules with silk-like domains as load-bearing elements in distal byssal threads. *Journal of Structural Biology* **2011,** *175*, 339-347.

30.   Waite, J. H., The Formation of Mussel Byssus: Anatomy of a Natural Manufacturing Process. In *Structure, Cellular Synthesis and Assembly of Biopolymers*, Case, S. T., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, 1992; pp 27-54.

31.   Pujol, J. P., Formation of byssus in common mussel (*Mytilus edulis*). *Nature* **1967,** *214*, 204-&.

32.   Vitellaro-Zuccarello, L., The collagen gland of *Mytilus galloprovincialis*: an ultrastructural and cytochemical study on secretory granules. *Journal of Ultrastructure Research* **1980,** *73*, 135-147.

33.   Bdolah, A.; Keller, P. J., Isolation of collagen granules from foot of sea mussel, *Mytilus californianus*. *Comparative biochemistry and physiology* **1976,** *55*, 171-174.

34.   Tamarin, A.; Keller, P. J., Ultrastructural study of byssal thread forming system in *Mytilus*. *Journal of Ultrastructure Research* **1972,** *40*, 401-416.

35.   Li, Y. F.; Suen, J. J. Y.; Prince, E.; Larin, E. M.; Klinkova, A.; Therien-Aubin, H.; Zhu, S. J.; Yang, B.; Helmy, A. S.; Lavrentovich, O. D.; Kumacheva, E., Colloidal cholesteric liquid crystal in spherical confinement. *Nature Communications* **2016,** *7*, 12520.

36.   Harrington, M. J.; Waite, J. H., Holdfast heroics: Comparing the molecular and mechanical properties of Mytilus californianus byssal threads. *Journal of Experimental Biology* **2007,** *210*, 4307-4318.

37.   Lucas, J. M.; Vaccaro, E.; Waite, J. H., A molecular, morphometric and mechanical comparison of the structural elements of byssus from Mytilus edulis and Mytilus galloprovincialis. *Journal of Experimental Biology* **2002,** *205*, 1807-1817.

38.   Klok, H. A.; Lecommandoux, S., Supramolecular materials via block copolymer self-assembly. *Advanced Materials* **2001,** *13*, 1217-1229.

39.   Wong, C. K.; Qiang, X.; Müller, A. H. E.; Gröschel, A. H., Self-Assembly of block copolymers into internally ordered microparticles. *Progress in Polymer Science* **2020,** *102*, 101211.

40.   Vigild, M. E.; Chu, C.; Sugiyama, M.; Chaffin, K. A.; Bates, F. S., Influence of shear on the alignment of a lamellae-forming pentablock copolymer. *Macromolecules* **2001,** *34*, 951-964.

41.   Priemel, T.; Palia, R.; Babych, M.; Thibodeaux, C. J.; Bourgault, S.; Harrington, M. J., Compartmentalized processing of catechols during mussel byssus fabrication determines the destiny of DOPA. *Proceedings of the National Academy of Sciences* **2020,** *117*, 7613.

42.   Bairati, A.; Zuccarello, L. V., The ultrastructure of the byssal apparatus of Mytilus galloprovincialis   IV. Observation by transmission elctron microscopy. *Cell Tiss. Res.* **1976,** *166*, 219-234.

43.   Picken, S. J.; Sikkema, D. J.; Boerstoel, H.; Dingemans, T. J.; van der Zwaag, S., Liquid crystal main-chain polymers for high-performance fibre applications. *Liquid Crystals* **2011,** *38*, 1591-1605.

44.   Snedeker, J. G.; Foolen, J., Tendon injury and repair - A perspective on the basic mechanisms of tendon disease and future clinical therapy. *Acta Biomater* **2017,** *63*, 18-36.

45.   Dimitriou, R.; Tsiridis, E.; Giannoudis, P. V., Current concepts of molecular aspects of bone healing. *Injury* **2005,** *36*, 1392-1404.

46.   Giraud-Guille, M.-M., Liquid crystallinity in condensed type I collagen solutions : A clue to the packing of collagen in extracellular matrices. **1992,** *224*, 861.

47.   Nyström, G.; Mezzenga, R., Liquid crystalline filamentous biological colloids: Analogies and differences. *Current Opinion in Colloid & Interface Science* **2018,** *38*, 30-44.

48.   Rey, A. D., Liquid crystal models of biological materials and processes. *Soft Matter* **2010,** *6*, 3402-3429.

49.   Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410*, 541-548.

50.   Nystrom, G.; Arcari, M.; Mezzenga, R., Confinement-induced liquid crystalline transitions in amyloid fibril cholesteric tactoids. *Nature Nanotechnology* **2018,** *13*, 330-336.

51.   Dogic, Z.; Fraden, S., Smectic phase in a colloidal suspension of semiflexible virus particles. *Physical Review Letters* **1997,** *78*, 2417-2420.

52.   Collings, P.; Hird, M.; Gray, G.; Goodby, J., *Introduction to Liquid Crystals*. CRC Press: London, 1997.

53.   Rezaei, N.; Lyons, A.; Forde, N. R., Environmentally controlled curvature of single collagen proteins. *Biophysical Journal* **2018,** *115*, 1457-1469.

54.   Collings, P. J., *Liquid Crystals: Nature's Delicate Phase of Matter*. Princeton University Press: Princeton, 2002

55.   Reinecke, A.; Brezesinski, G.; Harrington, M. J., pH-responsive self-organization of metal-binding protein motifs from biomolecular junctions in mussel byssus *Advanced Materials Interfaces* **2017,** *4*, 1600416.

56.   Schmidt, S.; Reinecke, A.; Wojcik, F.; Pussak, D.; Hartmann, L.; Harrington, M. J., Metal-mediated molecular self-healing in histidine-rich mussel peptides. *Biomacromolecules* **2014,** *15*, 1644-1652.

57.   Jehle, F.; Fratzl, P.; Harrington, M. J., Metal-tunable self-assembly of hierarchical structure in mussel-inspired peptide films. *ACS Nano* **2018,** *12*, 2160-2168.

58.   Trapaidze, A.; D'Antuono, M.; Fratzl, P.; Harrington, M. J., Exploring mussel byssus fabrication with peptide-polymer hybrids: Role of pH and metal coordination in self-assembly and mechanics of histidine-rich domains. *European Polymer Journal* **2018,** *109*, 229-236.

59.   Fullenkamp, D. E.; He, L.; Barrett, D. G.; Burghardt, W. R.; Messersmith, P. B., Mussel-inspired histidine-based transient network metal coordination hydrogels. *Macromolecules* **2013,** *46*, 1167–1174.

60.   Zechel, S.; Hager, M. D.; Priemel, T.; Harrington, M. J., Healing through histidine: Bioinspired pathways to self-healing polymers via imidazole‾metal coordination. *Biomimetics* **2019,** *4*, 20.

61.   Enke, M.; Bose, R. K.; Zechel, S.; Vitz, J.; Deubler, R.; Garcia, S. J.; van der Zwaag, S.; Schacher, F. H.; Hager, M. D.; Schubert, U. S., A translation of the structure of mussel byssal

threads into synthetic materials by the utilization of histidine-rich block copolymers. *Polymer Chemistry* **2018,** *9*, 3543-3551.

62.   Daly, A. C.; Riley, L.; Segura, T.; Burdick, J. A., Hydrogel microparticles for biomedical applications. *Nature Reviews Materials* **2020,** *5*, 20-43.

63.   Kretlow, J. D.; Klouda, L.; Mikos, A. G., Injectable matrices and scaffolds for drug delivery in tissue engineering. *Advanced Drug Delivery Reviews* **2007,** *59*, 263-273.

64.   Krull, A.; Buchholz, T.-O.; Jug, F., Noise2Void - Learning denoising from single noisy images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* **2019**, 2124-2132.

65.   Sauvola, J.; Pietikäinen, M., Adaptive document image binarization. *Pattern Recognition* **2000,** *33*, 225-236.

66.   Black, S. D.; Mould, D. R., Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Analytical Biochemistry* **1991,** *193*, 72-82.